

Claims

- [c1] 1. A method for providing access to data in a multi-threaded computing system, the method comprising:
providing a cache containing pages of data in memory of the multi-threaded computing system;
associating a latch with each page in the cache to regulate access to the page, the latch allowing multiple threads to share access to the page for read operations and a single thread to obtain exclusive access to the page for write operations;
in response to a request from a first thread to read a particular page, determining whether the particular page is in the cache without blocking access by other threads to pages in the cache;
if the particular page is in the cache, attempting to obtain the latch for purposes of reading the particular page; and
allowing the first thread to read the particular page unless a second thread has latched the particular page on an exclusive basis.
- [c2] 2. The method of claim 1, wherein the pages of data comprise database file pages.

- [c3] 3. The method of claim 1, wherein said providing step includes organizing the pages of data in an array.
- [c4] 4. The method of claim 1, wherein said providing step includes providing an index facilitating access to pages in the cache.
- [c5] 5. The method of claim 4, wherein said step of providing an index includes providing an index based upon a unique identifier assigned to each page in the cache.
- [c6] 6. The method of claim 1, wherein said providing step includes maintaining state information for each page in the cache.
- [c7] 7. The method of claim 6, wherein said step of maintaining state information for each page includes assigning a unique identifier to each page.
- [c8] 8. The method of claim 7, wherein said unique identifier comprises a page name.
- [c9] 9. The method of claim 7, wherein said determining step includes determining whether the particular page is in cache based, at least in part, upon said unique identifier.
- [c10] 10. The method of claim 6, wherein the state information maintained for each page includes the latch for regulat-

ing access to the page.

[c11] 11. The method of claim 1, wherein said determining step includes determining whether the particular page is in the cache without acquiring a mutual exclusion object (mutex) controlling access to pages in the cache.

[c12] 12. The method of claim 1, wherein said allowing step includes allowing the first thread to read the particular page concurrently with another thread reading the particular page.

[c13] 13. The method of claim 1, further comprising:
if the particular page is not in the cache, attempting to install a cache entry for the particular page.

[c14] 14. The method of claim 13, wherein said attempting to install step includes obtaining an unused cache entry.

[c15] 15. The method of claim 14, wherein said attempting to install step includes latching the unused cache entry on an exclusive basis.

[c16] 16. The method of claim 13, wherein said attempting to install step includes acquiring a mutual exclusion object (mutex) controlling access to a cache chain and attempting to install the unused cache entry in the cache chain.

[c17] 17. The method of claim 1, further comprising:

if a second thread has latched the particular page on an exclusive basis, attempting to prevent reuse of the particular page.

- [c18] 18. The method of claim 17, wherein said attempting to prevent reuse step includes determining if the particular page is in the cache.
- [c19] 19. The method of claim 17, wherein said attempting to prevent reuse step includes incrementing an indicator associated with the particular page so as to indicate the first thread is waiting for access to the particular page.
- [c20] 20. The method of claim 1, further comprising:
maintaining a list of reusable pages representing pages in the cache that are available for reuse.
- [c21] 21. The method of claim 20, wherein the list of reusable pages is structured as a double ended queue.
- [c22] 22. The method of claim 21, further comprising:
maintaining a gap between ends of the double ended queue.
- [c23] 23. The method of claim 21, further comprising:
identifying a page in the cache that can be reused; and
adding the identified page to one end of the double ended queue.

- [c24] 24. The method of claim 23, wherein said step of adding the identified page includes adding the identified page using thread-unsafe operations so as to provide improved system performance.
- [c25] 25. The method of claim 23, further comprising:
removing a page from the other end of the double ended queue when a page is needed.
- [c26] 26. The method of claim 25, wherein said step of removing the page includes removing the page using thread-unsafe operations so as to provide improved system performance.
- [c27] 27. A computer-readable medium having processor-executable instructions for performing the method of claim 1.
- [c28] 28. A downloadable set of processor-executable instructions for performing the method of claim 1.
- [c29] 29. A system for providing access to data in a multi-threaded computing system, the system comprising:
a cache containing pages of data in memory of the multi-threaded computing system;
a latch associated with each page in the cache to regulate access to the page, the latch allowing multiple

threads to share access to the page for read operations and a single thread to obtain exclusive access to the page for write operations; and
a cache manager for receiving a request for reading a particular page from a first thread, determining whether the particular page is in the cache without blocking access by other threads to pages in the cache, attempting to obtain the latch for reading the particular page if the particular page is in the cache; and allowing the first thread to read the particular page unless another thread has latched the particular page on an exclusive basis.

- [c30] 30. The system of claim 29, wherein the pages of data comprise database file pages.
- [c31] 31. The system of claim 29, wherein the cache includes an array holding the pages of data.
- [c32] 32. The system of claim 29, wherein the cache includes an index facilitating access to the pages in the cache.
- [c33] 33. The system of claim 32, wherein said index is based upon a unique identifier assigned to each page in the cache.
- [c34] 34. The system of claim 29, wherein the cache includes state information for each page in the cache.

- [c35] 35. The system of claim 34, wherein the state information includes a unique identifier for each page.
- [c36] 36. The system of claim 35, wherein said unique identifier comprises a page name.
- [c37] 37. The system of claim 35, wherein the cache manager determines whether the particular page is in cache based, at least in part, upon said unique identifier.
- [c38] 38. The system of claim 29, wherein the cache manager determines whether the particular page is in the cache without acquiring a mutual exclusion object (mutex) regulating access to pages in the cache.
- [c39] 39. The system of claim 29, wherein the cache manager allows the first thread to read the particular page concurrently with another thread reading the particular page.
- [c40] 40. The system of claim 29, wherein the cache manager attempts to install a cache entry for the particular page if the particular page is not found in cache.
- [c41] 41. The system of claim 29, wherein the cache manager attempts to prevent reuse of the particular page if another thread has latched the page on an exclusive basis.
- [c42] 42. The system of claim 29, further comprising:

a list of reusable pages representing pages in the cache that are available for reuse.

- [c43] 43. The system of claim 42, wherein the list of reusable pages is structured as a double ended queue.
- [c44] 44. The system of claim 43, wherein the cache manager identifies a page in the cache that can be reused and adds the identified page to one end of the double ended queue.
- [c45] 45. The system of claim 44, wherein cache manager adds the identified page using thread-unsafe operations so as to provide improved system performance.
- [c46] 46. The system of claim 44, wherein the cache manager removes a page from the other end of the double ended queue when a page is needed.
- [c47] 47. The system of claim 46, wherein the cache manager removes the page using thread-unsafe operations so as to provide improved system performance.